

CACHE CONTROL METHOD AND CACHE CONTROL SYSTEM

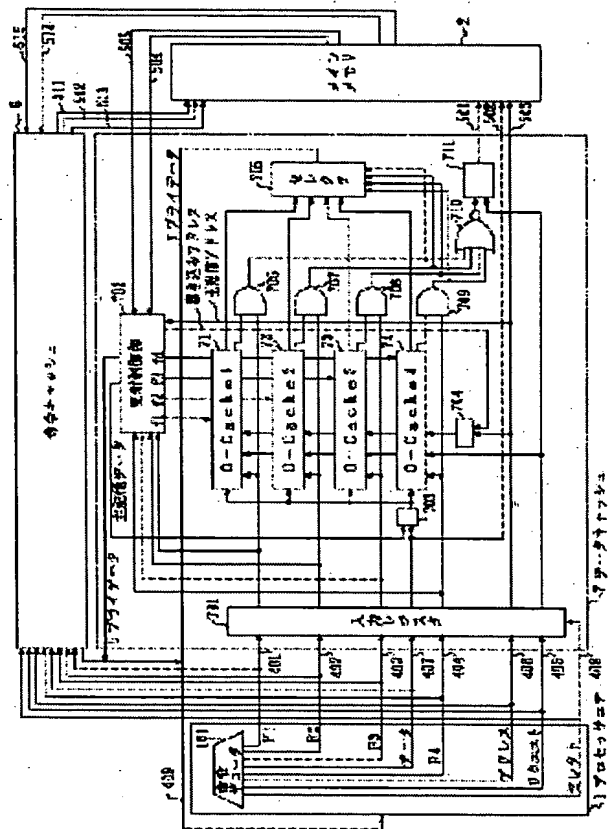
Patent number: JP2002116956
Publication date: 2002-04-19
Inventor: NAGANO TOMOAKI
Applicant: NEC CORP
Classification:
 - International: G06F12/08; G06F9/34; G06F9/46
 - european:
Application number: JP20000307121 20001006
Priority number(s):

Report a data error here

Abstract of JP2002116956

PROBLEM TO BE SOLVED: To directly select the optional number of caches to be actually used from the plural caches by a main memory access instruction.

SOLUTION: To the main memory access instructions such as a loading instruction executed in a processor core 1, cache selection flags F1-F4 in one- to-one correspondence with the caches 71-74 are added. At the time of executing the main memory access instructions, the cache selection flags F1-F4 are sent out to the respective caches 71-74. In the case of the loading instruction, the respective caches 71-74 judge a hit and an error based on a main memory address sent out from the processor core 1 and output judged results. AND circuits 706-709 AND the judged results and the cache selection flags F1-F4 and obtain a true hit judged result. In the case of the hit, cache data from the hit cache are selected in a selector 705 and returned to the processor core 1.



Data supplied from the esp@cenet database - Patent Abstracts of Japan

(18) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2002-116956

(P2002-116956A)

(43) 公開日 平成14年4月19日(2002.4.19)

(51) Int.Cl. ⁷	識別記号	FI	7-コード [*] (参考)	
G 0 6 F 12/08	5 2 3	G 0 6 F 12/08	5 2 3 B	5 B 0 0 5
	5 1 1		5 1 1 B	5 B 0 3 9
			5 1 1 E	5 B 0 9 8
	5 5 1		5 5 1	
	5 5 5		5 5 5	

審査請求 有 請求項の数14 O L (全 16 頁) 最終頁に続く

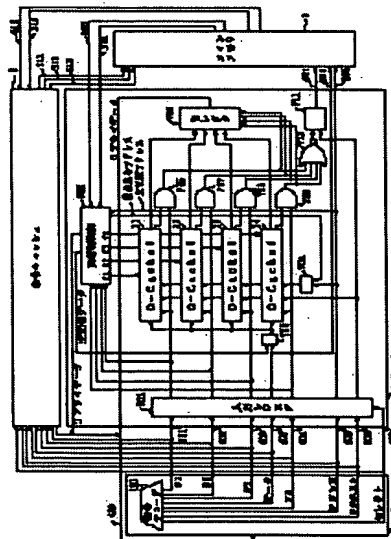
(21) 出願番号	特願2000-307121(P2000-307121)	(71) 出願人	000004237 日本電気株式会社 東京都港区芝五丁目7番1号
(22) 出願日	平成12年10月6日(2000.10.6)	(72) 発明者	長野 知明 東京都港区芝五丁目7番1号 日本電気株式会社内
		(74) 代理人	100088959 弁理士 境 廣巳
		Pターム(参考)	5B005 JJ13 KK12 LL01 LL11 MM02 MM03 TT02 5B033 AA04 AA15 DB01 DB12 DB14 5B098 AA03 GA02 GA04 GG03 GG14

(54) 【発明の名称】 キャッシュ制御方法及びキャッシュ制御システム

(57) 【要約】

【課題】 複数のキャッシュの中から実際に使用する任意個数のキャッシュを、主記憶アクセス命令によって直接選択し得るようにする。

【解決手段】 プロセッサコア1で実行されるロード命令等の主記憶アクセス命令には、キャッシュ71〜74に1対1に対応するキャッシュ選択フラグF1〜F4が付加されており、その主記憶アクセス命令の実行時、各キャッシュ71〜74にキャッシュ選択フラグF1〜F4が送出される。ロード命令の場合、各キャッシュ71〜74は、プロセッサコア1から送出された主記憶アドレスに基づきヒット、ミスの判定を行い、判定結果を出力する。アンド回路706〜709はその判定結果とキャッシュ選択フラグF1〜F4との論理積をとり、真のヒット判定結果を得る。ヒットした場合、ヒットしたキャッシュからのキャッシュデータをセレクト705で選択し、プロセッサコア1へ返送する。



【特許請求の範囲】

【請求項1】 キャッシュのセットに1対1に対応するキャッシュ選択フラグを有する主記憶アクセス命令によって主記憶データの写しを保持するキャッシュのセットを複数有するキャッシュメモリの中から、当該主記憶アクセス命令の実行に際して使用するキャッシュのセットを選択するキャッシュ制御方法。

【請求項2】 各キャッシュのセットが命令兼データ用のキャッシュのセットである請求項1記載のキャッシュ制御方法。

【請求項3】 命令キャッシュとデータキャッシュの少なくとも一方をそれぞれ複数のキャッシュのセットで構成し、命令キャッシュのアクセスにかかる主記憶アクセス命令の場合は命令キャッシュ中のセットをキャッシュ選択フラグによって選択し、データキャッシュのアクセスにかかる主記憶アクセス命令の場合はデータキャッシュ中のセットをキャッシュ選択フラグによって選択する請求項1記載のキャッシュ制御方法。

【請求項4】 命令キャッシュを構成するキャッシュのセット数をM、データキャッシュを構成するキャッシュのセット数をN、MとNの内の大きな方の値をMNとすると、主記憶アクセス命令にMN個のキャッシュ選択フラグを設け、命令キャッシュのアクセスにかかる主記憶アクセス命令の場合はMN個のキャッシュ選択フラグの内のN個のキャッシュ選択フラグを使用し、データキャッシュのアクセスにかかる主記憶アクセス命令の場合はMN個のキャッシュ選択フラグの内のM個のキャッシュ選択フラグを使用する請求項3記載のキャッシュ制御方法。

【請求項5】 命令キャッシュを構成するキャッシュのセット数をM、データキャッシュを構成するキャッシュのセット数をNとすると、主記憶アクセス命令にM個の命令用のキャッシュ選択フラグとN個のデータ用のキャッシュ選択フラグとを設け、命令キャッシュのアクセスにかかる主記憶アクセス命令の場合は命令用のキャッシュ選択フラグを使用し、データキャッシュのアクセスにかかる主記憶アクセス命令の場合はデータ用のキャッシュ選択フラグを使用する請求項3記載のキャッシュ制御方法。

【請求項6】 主記憶アクセス命令がロード命令のときは、キャッシュ選択フラグで選択したキャッシュのセットの何れかでキャッシュヒットした場合に限って当該ヒットしたデータをリプライデータとし、主記憶アクセス命令がストア命令のときは、キャッシュ選択フラグで選択したキャッシュのセットの何れかでキャッシュヒットした場合に限って当該ヒットしたデータを更新する請求項1、2、3、4または5記載のキャッシュ制御方法。

【請求項7】 主記憶アクセス命令がロード命令のときにキャッシュ選択フラグで選択した全てのキャッシュのセットでキャッシュミスした場合、主記憶から読み出し

た内容によって、前記キャッシュ選択フラグで選択したキャッシュのセットの何れか1つのセットを更新する請求項6記載のキャッシュ制御方法。

【請求項8】 キャッシュのセットに1対1に対応するキャッシュ選択フラグを有する主記憶アクセス命令を実行するプロセッサコアと、該プロセッサコアと主記憶との間に設けられたキャッシュメモリとを備え、該キャッシュメモリは、主記憶データの写しを保持するキャッシュのセットを複数有し、且つ、前記プロセッサコアから主記憶アクセス命令の実行時に出力されるキャッシュ選択フラグに基づいて前記複数のキャッシュのセットの中から当該主記憶アクセス命令の実行に際して使用するキャッシュのセットを選択するキャッシュコントローラを有するキャッシュ制御システム。

【請求項9】 各キャッシュのセットが命令兼データ用のキャッシュのセットである請求項8記載のキャッシュ制御システム。

【請求項10】 前記キャッシュメモリは、少なくとも一方が複数のキャッシュのセットで構成された命令キャッシュとデータキャッシュを備え、前記キャッシュコントローラは、命令キャッシュのアクセスにかかる主記憶アクセス命令の場合は命令キャッシュ中のセットをキャッシュ選択フラグによって選択し、データキャッシュのアクセスにかかる主記憶アクセス命令の場合はデータキャッシュ中のセットをキャッシュ選択フラグによって選択する請求項8記載のキャッシュ制御システム。

【請求項11】 命令キャッシュを構成するキャッシュのセット数をM、データキャッシュを構成するキャッシュのセット数をN、MとNの内の大きな方の値をMNとすると、主記憶アクセス命令にはMN個のキャッシュ選択フラグが設けられ、前記プロセッサコア及び前記キャッシュコントローラは、命令キャッシュのアクセスにかかる主記憶アクセス命令の場合はMN個のキャッシュ選択フラグの内のN個のキャッシュ選択フラグを使用し、データキャッシュのアクセスにかかる主記憶アクセス命令の場合はMN個のキャッシュ選択フラグの内のM個のキャッシュ選択フラグを使用する請求項10記載のキャッシュ制御システム。

【請求項12】 命令キャッシュを構成するキャッシュのセット数をM、データキャッシュを構成するキャッシュのセット数をNとすると、主記憶アクセス命令にはM個の命令用のキャッシュ選択フラグとN個のデータ用のキャッシュ選択フラグとが設けられ、前記プロセッサコア及び前記キャッシュコントローラは、命令キャッシュのアクセスにかかる主記憶アクセス命令の場合は命令用のキャッシュ選択フラグを使用し、データキャッシュのアクセスにかかる主記憶アクセス命令の場合はデータ用のキャッシュ選択フラグを使用する請求項10記載のキャッシュ制御システム。

【請求項13】 前記キャッシュコントローラは、主記

憶アクセス命令がロード命令のときは、キャッシュ選択フラグで選択したキャッシュのセットの何れかでキャッシュヒットした場合に限って当該ヒットしたデータをリプライデータとし、主記憶アクセス命令がストア命令のときは、キャッシュ選択フラグで選択したキャッシュのセットの何れかでキャッシュヒットした場合に限って当該ヒットしたデータを更新する請求項8、9、10、11または12記載のキャッシュ制御システム。

【請求項14】 前記キャッシュコントローラは、主記憶アクセス命令がロード命令のときにキャッシュ選択フラグで選択した全てのキャッシュのセットでキャッシュミスした場合、主記憶から読み出した内容によって、前記キャッシュ選択フラグで選択したキャッシュのセットの何れか1つのセットを更新する請求項13記載のキャッシュ制御システム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は主記憶データの写しを保持するキャッシュメモリに関し、より具体的には、複数のキャッシュを備え、使用するキャッシュをロード命令などの主記憶アクセス命令によって任意に選択することができるキャッシュ制御方法及びキャッシュ制御システムに関する。

【0002】

【従来の技術】 一般に現在の計算機システムは、プロセッサと主記憶との間に高速、小容量のキャッシュメモリを配置し、主記憶に置かれたデータ（命令も含む）へのアクセスの高速化を図っている。また複数のキャッシュを搭載し、目的に応じてキャッシュを使い分けることで、更なる高速化を図った技術が以下に述べるように幾つか提案されている。

【0003】 特開平2-18644号公報には、アプリケーション用のキャッシュと割り込み処理用のキャッシュとを設け、割り込み検出時には割り込み処理用のキャッシュを選択してオペレーティングシステムプログラムによる割り込み処理を実行し、その割り込み処理終了時に、割り込み直前に使用していたアプリケーション用のキャッシュを選択する技術が記載されている。同公報に記載された技術を以下、第1の従来技術と呼ぶ。

【0004】 特開昭51-267149号公報には、複数のキャッシュのうち、プログラムを実行するときのプロセッサの実行レベル情報に応じたキャッシュを選択することにより、最高位の実行レベルを持つオペレーティングシステムプログラムに1つのキャッシュを割り当て、他の複数存在するアプリケーションプログラムにはその他のキャッシュを割り当てるようにした技術が記載されている。同公報に記載された技術を以下、第2の従来技術と呼ぶ。

【0005】 特開平10-207770号公報には、命令キャッシュとデータキャッシュとを有するキャッシュ

メモリにおいて、特権モード用命令キャッシュとユーザモード用命令キャッシュとの2種類の命令キャッシュを設け、プロセッサの実行モードが特権モードか、ユーザモードかに応じて上記いずれかの命令キャッシュを選択し、また、グローバル変数用データキャッシュとローカル変数用データキャッシュとの2種類のデータキャッシュを設け、アドレスによりデータがグローバル変数かローカル変数かを判別し、上記いずれかのデータキャッシュを選択する技術が記載されている。また、同公報には、命令キャッシュ及びデータキャッシュをそれぞれ複数の単位ユニットで構成し、キャッシュ構造レジスタという特別なレジスタの設定値に応じて、各用途に使うキャッシュのユニット数を動的に変更する考えが提案されている。以下、同公報に記載された技術を第3の従来技術と呼ぶ。

【0006】 特開平5-143451号公報には、複数のキャッシュメモリの一方を主記憶データの先取り用キャッシュとして、他方を一般的なキャッシュとして使い分ける技術が示されている。また、キャッシュを選択する手法として、ロード命令などの主記憶アクセス命令に付加したキャッシュ番号によって、複数存在するキャッシュの中の何れか1つのキャッシュを選択する手法が提案されている。以下、同公報に記載された技術を第4の従来技術と呼ぶ。

【0007】

【発明が解決しようとする課題】 一般に複数のプロセスを並行して実行する計算機システムでは、実行されるプロセスの特性によってキャッシュの有効性には様々な違いが発生する。第1乃至第3の従来技術は主に、オペレーティングシステムにかかるプロセスとユーザアプリケーションプロセスにかかるプロセスとの特性の差異に着目し、双方のプロセスでキャッシュを使い分けることで、キャッシュの有効性を高めている。

【0008】 しかしながら、ユーザアプリケーションプロセスであっても、その特性は様々である。例えば、大規模な容量のメモリをランダムにアクセスするようなプロセスでは、キャッシュミスが頻発し、キャッシュの中身が書き換わるのみで効率ほとんど上がらないが、データのローカライズが適正に行われたプロセスではキャッシュの有効性は大きい。従って、キャッシュが有効なプロセスと有効ではないプロセスが同時に実行される場合、それらのプロセスに同じキャッシュを割り当てると、キャッシュを有効活用できないプロセスのメモリアクセスが、キャッシュを有効活用できるプロセスの使用しているキャッシュ領域を更新し続け、結果としてキャッシュを有効活用できるはずのプロセスのキャッシュアクセス効率まで低下させてしまう。また、共にキャッシュを有効にアクセスできるプロセス同士の動作でも、例えばセットアソシアティブ型のキャッシュでは、同一のインデックスを持ち、違うタグアドレスを持つアクセス

がプロセス毎に発生することにより、お互いのプロセスのキャッシュ領域を侵害し、キャッシュの使用効率が下がる可能性も大きい。

【0009】このようなことから、キャッシュの選択は、オペレーティングプログラムプロセスとユーザアプリケーションプロセス間、異なる実行レベルのプロセス間といった大雑把な単位ではなく、個々のプロセス単位で行えることが望ましい。しかし、第1乃至第3の従来技術では、割り込みやプロセッサの実行レベルに基づいてキャッシュを切り替えるため、プロセス単位でキャッシュを選択することは不可能である。

【0010】これに対して、ロード命令などの主記憶アクセス命令に付加したキャッシュ番号によってキャッシュ選択を行う第4の従来技術を使えば、プロセス毎に、そのプロセスで使用する主記憶アクセス命令中のキャッシュ番号を制御できるので、プロセス単位でのキャッシュ選択が可能になる。

【0011】しかしながら、第4の従来技術のようにキャッシュ番号によってキャッシュを選択する構成では、選択できるキャッシュは、指定したキャッシュ番号に対応する1つのキャッシュに限られる。例えばキャッシュ番号a、b、c、dのキャッシュA、B、C、Dがある場合、その何れか1つのキャッシュは選択できるが、キャッシュAとキャッシュBのように異なるキャッシュ番号の複数のキャッシュを選択することはできない。つまり、使用するキャッシュの構成を自由に編成できないため、実行中のプロセスに適した容量のキャッシュを用いるといった制御が困難である。キャッシュ構成レジスタの設定値に応じてキャッシュの構成を編成する第3の従来技術と組み合わせることも考えられるが、キャッシュ構成レジスタへの設定に主記憶アクセス命令とは別の1命令が必要になるため、オーバーヘッドが大きくなってしまふ。

【0012】更に、第1乃至第4の従来技術においては、少なくとも1つのキャッシュは必ず選択されるようになっている為、或るプロセスについてはキャッシュの使用を禁止するといった制御は行えない。

【0013】そこで本発明の目的は、複数のキャッシュを備えたキャッシュメモリにおいて、複数のキャッシュの中から実際に使用する任意個数のキャッシュを、主記憶アクセス命令によって直接選択し得るようにすることにある。

【0014】

【課題を解決するための手段】本発明のキャッシュ制御方法は、キャッシュのセットに1対1に対応するキャッシュ選択フラグを有する主記憶アクセス命令によって、主記憶データの写しを保持するキャッシュのセットを複数有するキャッシュメモリの中から、当該主記憶アクセス命令の実行に際して使用するキャッシュのセットを選択する。

【0015】また、本発明のキャッシュ制御システムは、キャッシュのセットに1対1に対応するキャッシュ選択フラグを有する主記憶アクセス命令を実行するプロセッサコアと、該プロセッサコアと主記憶との間に設けられたキャッシュメモリとを備え、該キャッシュメモリは、主記憶データの写しを保持するキャッシュのセットを複数有し、且つ、前記プロセッサコアから主記憶アクセス命令の実行時に出力されるキャッシュ選択フラグに基づいて前記複数のキャッシュのセットの中から当該主記憶アクセス命令の実行に際して使用するキャッシュのセットを選択するキャッシュコントローラを有する。

【0016】上述のように構成された本発明のキャッシュ制御方法及びキャッシュ制御システムにあっては、キャッシュのセットに1対1に対応するキャッシュ選択フラグの状態に応じてキャッシュのセットが選択されるので、キャッシュ選択フラグの設定の仕方によって、複数のキャッシュのセットの中から1つのセットを選択する以外に、2つ以上のセットを選択したり、全てのセットを選択することも可能であり、またセットを1つも選択しないといったことも可能な柔軟な制御が可能となる。更に、任意数のセットの選択が主記憶アクセス命令毎に行えるため、プロセス毎に、そのプロセスで使用する主記憶アクセス命令中のキャッシュ選択フラグを調整することで、プロセス単位でのキャッシュ選択が可能になり、且つ実行中のプロセスに適した容量のキャッシュを用いるといったキャッシュ構成をオーバーヘッド無しに実施することが可能となる。

【0017】本発明は、命令用とデータ用とで同じキャッシュを使うキャッシュメモリに対しても適用でき、双方で別々のキャッシュを使うキャッシュメモリに対しても適用できる。前者のキャッシュメモリでは、命令兼データ用のキャッシュのセットが複数存在し、主記憶アクセス命令中のキャッシュ選択フラグに応じて、使用するキャッシュのセットが選択される。

【0018】後者のキャッシュメモリでは、命令キャッシュとデータキャッシュの少なくとも一方が複数のキャッシュのセットで構成される。そして、命令キャッシュ及びデータキャッシュの中から、実際に使用するキャッシュのセットが主記憶アクセス命令中のキャッシュ選択フラグに応じて選択される。

【0019】命令キャッシュとデータキャッシュの少なくとも一方が複数のキャッシュのセットで構成されている場合、先ず、主記憶アクセス命令が命令キャッシュへのアクセスか、データキャッシュへのアクセスかに応じて命令キャッシュか、データキャッシュを選択し、次に選択した命令キャッシュ、データキャッシュの中から、実際に使用するキャッシュのセットを主記憶アクセス命令中のキャッシュ選択フラグに応じて選択する。この場合、命令キャッシュ用とデータキャッシュ用とで別々のキャッシュ選択フラグを設けるようにしても良いし、双

方で共通のキャッシュ選択フラグを使うようにしても良い。

【0020】キャッシュ選択フラグを命令キャッシュ用とデータキャッシュ用とで別々に設ける場合には、命令キャッシュを構成するキャッシュのセット数をM、データキャッシュを構成するキャッシュのセット数をNとすると、主記憶アクセス命令にM個の命令用のキャッシュ選択フラグとN個のデータ用のキャッシュ選択フラグとが設けられ、命令キャッシュのアクセスにかかる主記憶アクセス命令の場合は命令用のキャッシュ選択フラグが使用され、データキャッシュのアクセスにかかる主記憶アクセス命令の場合はデータ用のキャッシュ選択フラグが使用される。

【0021】キャッシュ選択フラグを命令キャッシュ用とデータキャッシュ用とで共通化する場合、命令キャッシュを構成するキャッシュのセット数をM、データキャッシュを構成するキャッシュのセット数をN、MとNの内の大きな方の値をMNとすると、主記憶アクセス命令にMN個のキャッシュ選択フラグが設けられ、命令キャッシュのアクセスにかかる主記憶アクセス命令の場合はMN個のキャッシュ選択フラグの内のN個のキャッシュ選択フラグが使用され、データキャッシュのアクセスにかかる主記憶アクセス命令の場合はMN個のキャッシュ選択フラグの内のM個のキャッシュ選択フラグが使用される。

【0022】また本発明のキャッシュ制御方法及びキャッシュ制御システムにおいては、主記憶アクセス命令がロード命令のときは、キャッシュ選択フラグで選択したキャッシュのセットの何れかでキャッシュヒットした場合に限って当該ヒットしたデータをリプライデータとし、主記憶アクセス命令がストア命令のときは、キャッシュ選択フラグで選択したキャッシュのセットの何れかでキャッシュヒットした場合に限って当該ヒットしたデータを更新する。更に、主記憶アクセス命令がロード命令のときにキャッシュ選択フラグで選択した全てのキャッシュのセットでキャッシュミスした場合、主記憶から読み出した内容によって、前記キャッシュ選択フラグで選択したキャッシュのセットの何れか1つのセットを更新する。

【0023】

【発明の実施の形態】次に本発明の実施の形態の例について図面を参照して詳細に説明する。

【0024】図1は本発明を適用した計算機システムの一例を示すブロック図である。この例の計算機システムは、プロセッサコア1と主記憶（メインメモリ）2との間にキャッシュメモリ3を備え、プロセッサコア1から主記憶2へのアクセスはキャッシュメモリ3を介して行う。なお、キャッシュメモリ3は、プロセッサコア1と共にチップにLSI化されていても良く、別LSIで構成されていても良い。

【0025】キャッシュメモリ3は、プロセッサコア1とはプロセッサバス4によって接続され、主記憶2とは主記憶バス5によって接続されており、命令キャッシュ6とデータキャッシュ7とを含んでいる。命令キャッシュ6は、主記憶2の内容のうち命令の写しを保持するもので、4個のキャッシュ61～64と、これらのキャッシュ61～64を制御するキャッシュコントローラ65とを含む。また、データキャッシュ7は、主記憶2の内容のうちデータの写しを保持するもので、命令キャッシュ6と同様に4個のキャッシュ71～74と、これらのキャッシュ71～74を制御するキャッシュコントローラ75とを含む。キャッシュコントローラ65、75は、プロセッサバス4を介してプロセッサコア1に接続され、主記憶バス5を介して主記憶2に接続される。

【0026】本実施形態では、キャッシュ61とキャッシュ71とがセット1のキャッシュ、キャッシュ62とキャッシュ72とがセット2のキャッシュ、キャッシュ63とキャッシュ73とがセット3のキャッシュ、キャッシュ64とキャッシュ74とがセット4のキャッシュを構成している。つまり、本実施形態のキャッシュメモリ3は、命令用とデータ用とにそれぞれ4セットずつキャッシュを備えている。

【0027】個々のキャッシュ61～64、71～74は全て同じ構成であり、主記憶データの一部の写しをブロック単位で格納するデータアレイと、データアレイに格納されているデータブロックのアドレスを格納するアドレスアレイと、キャッシュヒット/ミスの判定を行う比較器などから構成される。キャッシュの構成としては、例えば1WAY、2WAY、4WAY等のnWAY、セットアソシアティブ型のキャッシュが使われる。勿論、ダイレクトマッピング型のキャッシュ等、他の形式のキャッシュを使用することもできる。

【0028】本実施形態では、命令キャッシュ6及びデータキャッシュ7のそれぞれに存在する4セットのキャッシュの中から任意のキャッシュを主記憶アクセス命令によって選択して使用する。このために、主記憶アクセス命令には、キャッシュのセットに1対1に対応するキャッシュ選択フラグが付加されている。

【0029】本例の計算機システムで使われる主記憶アクセス命令のうち、一般的なロード命令とストア命令についてのフォーマット例を図2に示す。本発明に関連する部分は、ロード命令およびストア命令の双方に付加された図中のFの部分である。この部分が、ロード命令およびストア命令において使用するキャッシュを選択するキャッシュ選択フラグである。なお、周知のようにロード命令及びストア命令は、オペコード（オペレーションコード部）とアドレス部とを含む。ロード命令のオペコードには、命令フェッチとデータフェッチとの2種類がある。アドレス部では、主記憶アドレスやレジスタの指定情報など、命令の対象となるオペランドを示す。

【0030】図3にキャッシュ選択フラグFの内容例を示す。本例では、キャッシュ選択フラグFは、F1、F2、F3、F4の合計4つのフラグから構成される。フラグF1はセット1のキャッシュ61、71に、フラグF2はセット2のキャッシュ62、72に、フラグF3はセット3のキャッシュ63、73に、フラグF4はセット4のキャッシュ64、74にそれぞれ1対1に対応している。各フラグF1～F4のそれぞれは、発行された主記憶アクセス命令において、対応するセットのキャッシュを用いる場合には論理“1”に、対応するセットのキャッシュを用いない場合には論理“0”にそれぞれセットされる。

【0031】論理“1”にしておくフラグの位置および数は任意である。例えば、或るプロセスで使用するロード命令及びストア命令では、フラグF1のみ論理“1”とし、別のプロセスで使用するロード命令及びストア命令では、フラグF2～F4を論理“1”とすることができる。このため、キャッシュ選択フラグFにより、使用するキャッシュの構成を自由に編成することができ、実行中のプロセスに適した容量のキャッシュを用いることが可能である。また、或るプロセスで使用するロード命令及びストア命令では、全てのフラグF1～F4を論理“0”にしており、キャッシュの使用を禁止することもできる。これにより、大容量のメモリをランダムにアクセスするようなキャッシュが効果的に働かないプロセスを実行する際に、他のプロセスが使用しているキャッシュの内容を更新しないようにすることが可能となる。

【0032】次に、図1の実施形態の概略動作を説明する。先ず、ロード命令実行時の動作を説明する。

【0033】プロセッサコア1においてロード命令の実行が開始されると、そのオペコードのデコード結果に応じて、命令フェッチの場合には命令キャッシュ6が、データフェッチの場合にはデータキャッシュ7がそれぞれ選択され、ロードリクエスト、主記憶アドレス及び当該ロード命令中のフラグF1～F4がプロセッサバス4を通じて、選択されたキャッシュに送られる。以降の動作は命令キャッシュ6とデータキャッシュ7とで同じなので、データキャッシュ7を例に説明する。

【0034】データキャッシュ7では、対応するフラグF1～F4が論理“1”となっているセットのキャッシュ71～74でキャッシュヒットした場合に限って、ヒットしたデータがキャッシュコントローラ75からプロセッサバス4を通じてプロセッサコア1に返却される。この場合、キャッシュコントローラ75が、対応するフラグF1～F4が論理“1”となっているセットのキャッシュ71～74のみを動作させてヒット判定を行わせ、ヒットしたキャッシュが存在した場合にそのヒットデータをプロセッサコア1に返却するよう構成しても良いし、後述する実施例のように全てのセットのキャッシュ71～74を動作させてヒット判定を行わせ、キャッ

シュコントローラ75が、対応するフラグF1～F4が論理“1”となっている何れか1つのセットのキャッシュでヒットしたか否かを判定して、そのヒットデータをプロセッサコア1に返却するよう構成しても良い。

【0035】他方、対応するフラグF1～F4が論理“1”となっている全てのキャッシュでミスするか、フラグF1～F4の全てが論理“0”の場合には、キャッシュコントローラ75から主記憶2に対して主記憶バス5を通じてロードリクエストが送出される。その後、主記憶2から主記憶バス5を通じて、キャッシュミスしたデータを含む1ブロック分の主記憶データがデータキャッシュ7に返されてくると、キャッシュコントローラ75はその主記憶データの中からキャッシュミスしたデータを抽出してプロセッサバス4を通じてプロセッサコア1へ返却すると共に、フラグF1～F4の少なくとも1つが論理“1”である場合には、主記憶2から返された1ブロック分の主記憶データを、対応するフラグF1～F4が論理“1”となっている何れか1つのセットのキャッシュ71～74に書き込むことでキャッシュを更新する。

【0036】次にストア命令実行時の動作を説明する。プロセッサコア1においてストア命令の実行が開始されると、ストア命令はデータキャッシュ7に対してのみ行われるので、ストアリクエスト、主記憶アドレス、ストアデータ及び当該ストア命令中のフラグF1～F4がプロセッサバス4を通じてデータキャッシュ7に送られる。データキャッシュ7では、対応するフラグF1～F4が論理“1”となっているセットのキャッシュ71～74でキャッシュヒットした場合に限って、ヒットしたデータをプロセッサコア1から送られてきたストアデータで更新する。また、プロセッサコア1から送出されたストアリクエスト、主記憶アドレス及びストアデータはキャッシュコントローラ75から主記憶バス5を通じて主記憶2に送られ、主記憶2上の該当するアドレスのデータが当該ストアデータで上書きされる。つまり、本実施形態はストアスルー方式を採用している。

【0037】次に本実施形態の実施例について説明する。

【0038】図4は図1の計算機システムの構成をより具体化した実施例のブロック図である。プロセッサコア1に内蔵の命令デコーダ101で主記憶アクセス命令がデコードされると、その主記憶アクセス命令に含まれていたフラグF1～F4が結線401～404に、その主記憶アクセス命令のリクエスト種別（ロードか、ストアか）が結線405に、その主記憶アクセス命令でアクセスする対象となる主記憶アドレスが結線406に、その主記憶アドレス命令がストア命令の場合にはストアデータが結線407に、命令キャッシュ6及びデータキャッシュ7の何れか一方を選択するセレクト信号が結線408にそれぞれ出力される。プロセッサコア1と命令キャ

ッシュ6及びデータキャッシュ7とは、このような結線401~408と、リプライデータを伝送する結線409とによって相互に接続されている。

【0039】また、命令キャッシュ6及びデータキャッシュ7と主記憶2とは、リクエストを伝送する結線501、511、ストアデータを伝送する結線502、512、主記憶アドレスを伝送する結線503、513、主記憶データを伝送する結線504、514、結線504、514を通じて主記憶2からキャッシュ6、7に伝送する主記憶データのアドレス等を含む主記憶データ制御信号を伝送する結線505、515によって相互に接続される。

【0040】命令キャッシュ6とデータキャッシュ7とは、その内部構成が基本的に同じであるため、図4ではデータキャッシュ7のみ内部の構成を示し、命令キャッシュ6の内部構成は図示を省略している。データキャッシュ7内の構成のうち、キャッシュ71~74を除いた部分がキャッシュコントローラ75を構成する。

【0041】キャッシュコントローラ75は、入力レジスタ701、更新制御部702、セレクト703~705、アンド回路706~709、ノア回路710、リクエスト発行部711を含んでいる。

【0042】入力レジスタ701は、結線408上のセレクト信号によって自キャッシュが選択された場合に、結線401~404上のフラグF1~F4、結線405上の主記憶リクエスト、結線406上の主記憶アドレス、結線407上のストアデータを取り込み、各部に供給する。即ち、フラグF1~F4は、対応するキャッシュ71~74及びアンド回路706~709へ供給すると共に更新制御部702へ供給する。主記憶リクエスト種別は、各キャッシュ71~74とリクエスト発行部711へ供給する。主記憶アドレスは、セレクト704を通じて各キャッシュ71~74へ供給すると共に更新制御部702へ供給し、更に結線503を通じて主記憶2へ供給する。ストアデータは、セレクト703を通じて各キャッシュ71~74へ供給すると共に結線502を通じて主記憶2へ供給する。

【0043】アンド回路706~709は、各キャッシュ71~74から出力されるヒット/ミス判定信号（ヒット時に論理“1”、ミス時に論理“0”）と対応するフラグF1~F4との論理積信号をノア回路710及びセレクト705へ出力する。セレクト705は、各キャッシュ71~74からヒット時に出力されるキャッシュデータのうち、アンド回路706~709の出力が論理“1”になっているキャッシュのキャッシュデータを選択し、リプライデータとして結線409を通じてプロセッサコア1へ返却する。ノア回路710は、全てのアンド回路706~709の出力が論理“0”のときに、リクエスト発行部711へロードリクエストの発行を要求する信号を送出する。リクエスト発行部711は、この

ロードリクエストの発行要求にตอบสนองして、結線501を通じて主記憶2へロードリクエストを発行する。また、リクエスト発行部711は、入力レジスタ701から伝送されたリクエストがストアリクエストであれば、結線501を通じて主記憶2へストアリクエストを発行する。

【0044】更新制御部702は、結線504及び505を通じて主記憶2から1ブロック分の主記憶データ及び主記憶データ制御信号が送られてくると、入力レジスタ701から供給されている主記憶アドレス（ミスアドレス）のデータを1ブロック分の主記憶データから抽出して結線409を通じてリプライデータとしてプロセッサコア1へ返却する。同時に、入力レジスタ701から伝送されているフラグF1~F4のうち、論理“1”のフラグに対応する何れか1つのキャッシュ71~74を選択し、その選択したキャッシュに対して更新要求11~14を送出すると共に更新アドレスを送出することにより、キャッシュの更新制御を行う。

【0045】以下、図4を参照して本実施例におけるキャッシュ制御の動作を説明する。先ず、ロード命令実行時の動作を説明する。

【0046】プロセッサコア1においてロード命令が実行されると、そのオペコードのデコード結果に応じて、命令フェッチの場合には結線408上のセレクト信号によって命令キャッシュ6が選択され、データフェッチの場合には同セレクト信号によってデータキャッシュ7が選択される。以降の動作は命令キャッシュ6とデータキャッシュ7とで同じなので、データキャッシュ7を例に説明する。

【0047】データキャッシュ7の入力レジスタ701は、結線408のセレクト信号によって自キャッシュが選択されると、結線401~404上のフラグF1~F4を各キャッシュ71~74、アンド回路706~709及び更新制御部702へ、結線405上のリクエスト種別を各キャッシュ71~74及びリクエスト発行部711へ、それぞれ供給する。また、セレクト704を入力レジスタ701側に切り替えて、結線406上の主記憶アドレスをセレクト704を通じて各キャッシュ71~74へ供給すると共に、更新制御部702及び結線502に供給する。各キャッシュ71~74は、供給された主記憶アドレスに基づいてヒット、ミスの判定を実施し、ヒットした場合は論理“1”のヒット/ミス判定信号をアンド回路706~709へ出力すると同時にヒットデータをセレクト705へ出力する。他方、ミスした場合は論理“0”のヒット/ミス判定信号をアンド回路706~709へ出力する。アンド回路706~709は、対応するキャッシュ71~74のヒット/ミス判定信号と対応するフラグF1~F4との論理積をとり、それを真のヒット/ミス判定信号として出力する。つまり、通常はキャッシュ71~74から出力されるヒット

ノミス判定信号によってヒット、ミスが確定するわけだが、本実施例のデータキャッシュ7の場合、キャッシュ71~74自体はヒットと判定し、データを出力している場合でも、該当するフラグF1~F4の値が論理"0"であった場合はヒットとはならない。

【0048】アンド回路706~709から出力される真のヒット/ミス判定信号にヒットを示す論理"1"となるものがあつた場合、該当するキャッシュからは有効なデータが出力されていると判断し、セレクト705は、ヒットしたキャッシュからのキャッシュデータを選択し、リプライデータとして結線409を通じてプロセッサコア1へ返送する。

【0049】他方、アンド回路706~709から出力される真のヒット/ミス判定信号にヒットを示す論理"1"となるものがなかった場合、ノア回路710の出力が論理"1"となり、リクエスト発行部711は、結線501を通じて主記憶2に対してメモリアクセスリクエスト（ロードリクエスト）を発行する。このときの主記憶アドレスは結線503に供給されているアドレスが使用される。

【0050】主記憶2では、メモリアクセスリクエストに従って該当するデータを含む1ブロック分の主記憶データを読み出し、そのアドレスを含む主記憶データ制御信号と共に結線504、505を通じて、リクエスト元のデータキャッシュ7に送出する。更新制御部702は、結線504を通じて主記憶2から送られてくる1ブロック分の主記憶データ中から、入力レジスタ701より供給されているミスヒットした主記憶アドレスに該当するデータを抽出してリプライデータとして結線409を通じてプロセッサコア1へ返送する。また、入力レジスタ701から供給されているフラグF1~F4の中に論理"1"となるフラグが1つ以上存在すれば、任意のアルゴリズムに従って論理"1"のフラグの中から1つのフラグを選択し、この選択したフラグに対応する何れか1つのキャッシュ71~74に対して、キャッシュの更新要求11~14を送出すると共に、主記憶2から送られてきた1ブロック分の主記憶データをセレクト703を通じてキャッシュ71~74に伝達し、且つ、セレクト703を更新制御部702側に切り替えて、1ブロック分の主記憶データの書き込みアドレスをキャッシュ71~74に伝達する。キャッシュ71~74のうち、更新要求11~14によってキャッシュの更新が要求された何れか1つのキャッシュは、先ずセレクト704を通じて入力されている主記憶アドレスに基づいてアドレスアレイを更新し、次いで、セレクト703を通じて伝達される1ブロック分の主記憶データを、セレクト704を通じて伝達される書き込みアドレスに書き込むことで、データアレイを更新する。

【0051】次に、ストア命令実行時の動作を説明する。プロセッサコア1においてストア命令が発行される

と、結線408上のセレクト信号によってデータキャッシュ7が選択され、データキャッシュ7の入力レジスタ701は、結線401~404上のフラグF1~F4を各キャッシュ71~74、アンド回路706~709及び更新制御部702へ、結線405上のリクエスト種別を各キャッシュ71~74及びリクエスト発行部711へ、それぞれ供給する。また、セレクト704を入力レジスタ701側に切り替えて、結線406上の主記憶アドレスをセレクト704を通じて各キャッシュ71~74へ供給すると共に、更新制御部702及び結線502に供給する。更に、セレクト703を入力レジスタ701側に切り替えて、結線407上のストアデータをセレクト703を通じて各キャッシュ71~74へ供給すると共に、結線502を通じて主記憶2へ出力する。

【0052】各キャッシュ71~74は、供給された主記憶アドレスに基づいてヒット、ミスの判定を実施し、ヒットした場合であつて且つ自キャッシュに対応するフラグF1~F4が論理"1"の場合に限って、自キャッシュのデータを更新する。即ち、セレクト703を通じて伝達されたストアデータで、セレクト704を通じて伝達された主記憶アドレスに該当するキャッシュライン上のデータを更新する。また、リクエスト発行部711は、キャッシュのヒット、ミスにかかわらず、入力レジスタ701から伝達されたリクエスト種別（ストア）に基づき、結線501を通じて主記憶2に対してストアリクエストを発行する。このときの主記憶アドレスは結線503上に、ストアデータは結線502上にそれぞれ出力されている。

【0053】次に、各セットのキャッシュ71~74及び更新制御部702の構成例について説明する。

【0054】図5はキャッシュ71の構成例を示すブロック図である。この例のキャッシュ71は1WAYセットアソシアティブ構成のキャッシュであり、主記憶データの一部の写しをブロック単位で格納するデータアレイ（DA）711と、データアレイ711に格納されているデータブロックのアドレスを格納するアドレスアレイ（AA）712と、ヒット/ミスの判定を行うヒット判定器713と、制御部714と、アンド回路715とを主要部として含んでいる。主記憶アクセス命令で指示する主記憶アドレスは、その上位より順にタグアドレス、インデックスアドレス、ブロックアドレスに論理的に分けられ、アドレスアレイ712にはタグアドレスが登録される。

【0055】次に図5のキャッシュ71の動作を説明する。先ずロード命令時の動作を説明する。ロード命令の場合、ロード命令で指示された主記憶アドレス中のインデックスアドレスによりアドレスアレイ712の該当エントリがアクセスされ、ヒット判定器713により、そのエントリに記憶されているタグアドレスとロード命令で指示された主記憶アドレス中のタグアドレスとの比較

およびそのエントリ中の有効性指示ビットの検査が行われる。双方のタグアドレスが一致し且つ有効性指示ビットが有効を示しているときヒット、それ以外はミスとなる。ヒットであればヒット信号が、ミスであればミス信号がヒット判定器713から制御部714へ出力される。

【0056】制御部714は、リクエスト種別がロードであり且つヒット判定器713からヒット信号が入力されると、ヒットを示す論理“1”のヒット/ミス判定信号をアンド回路706へ出力すると共に、データアレイ711へ読み出しイネーブル信号を出力する。データアレイ711は、ロード命令で指示された主記憶アドレス中のインデックスアドレスで特定されるエントリ中の、そのブロックアドレスで特定されるデータを読み出し、セクタ705へ出力する。他方、制御部714は、リクエスト種別がロードであり且つヒット判定器713からミス信号が入力されると、ミスを示す論理“0”のヒット/ミス判定信号をアンド回路706へ出力する。

【0057】次に、キャッシュ更新時の動作を説明する。制御部714は、更新制御部702から更新要求が入力されると、先ずアドレスアレイ712へ書き込みイネーブル信号を出力する。これに応じてアドレスアレイ712は、ロード命令で指示された主記憶アドレス（ミスアドレス）中のタグアドレスをそのインデックスアドレスで特定されるエントリに書き込む。次に制御部714は、アンド回路715に書き込みイネーブル信号を出力する。制御部714が当該キャッシュ71に更新要求を出すときは当該キャッシュ71に対応するフラグF1は必ず論理“1”になっているため、アンド回路715からデータアレイ711へ書き込みイネーブル信号が出力される。データアレイ711は、更新制御部702からセクタ703を通じて送られてくる1ブロック分の主記憶データを、同じく更新制御部702からセクタ704を通じて送られてくる当該主記憶データのアドレス中のインデックスアドレスで特定されるエントリに書き込んでいく。

【0058】次に、ストア命令実行時の動作を説明する。ストア命令の場合、ストア命令で指示された主記憶アドレス中のインデックスアドレスによりアドレスアレイ712の該当エントリがアクセスされ、ヒット判定器713により、そのエントリに記憶されているタグアドレスとストア命令で指示された主記憶アドレス中のタグアドレスとの比較およびそのエントリ中の有効性指示ビットの検査が行われる。双方のタグアドレスが一致し且つ有効性指示ビットが有効を示しているときヒット、それ以外はミスとなる。ヒットであればヒット信号が、ミスであればミス信号がヒット判定器713から制御部714へ出力される。

【0059】制御部714は、リクエスト種別がストアであり且つヒット判定器713からヒット信号が入力さ

れると、アンド回路715へ書き込みイネーブル信号を出力する。アンド回路715は、自キャッシュ71に対応するフラグF1が論理“1”なら書き込みイネーブル信号をデータアレイ711へ伝達するが、フラグF1が論理“0”であれば伝達しない。データアレイ711は、書き込みイネーブル信号を受けると、ストア命令の主記憶アドレス中のインデックスアドレスで特定されるエントリにおけるブロックアドレスで特定されるデータを、ストアデータで書き換える。

【0060】図5ではキャッシュ71の構成例を示したが、データキャッシュ7の他のキャッシュ72~74及び命令キャッシュ6のキャッシュ61~64も同様に構成される。また、図5では1WAYセットアソシアティブ型のキャッシュを示したが、2WAY、4WAY等のnWAYセットアソシアティブ型のキャッシュを使用することもできる。

【0061】図6は更新制御部702の構成例を示すブロック図であり、制御部7021と選択部7022とで構成されている。制御部7021は、主記憶2から結線504及び505を通じて1ブロック分の主記憶データ及びそのアドレス等を含む主記憶データ制御信号が送られてくると、入力レジスタ701から入力されている主記憶アドレス（ミスアドレス）に基づいて1ブロック分の主記憶データから該当するデータを読み出して結線409へ出力する。また、更新するキャッシュを選択するために選択部7022を起動し、1ブロック分の主記憶データをセクタ703を通じてキャッシュ71~74へ送出すると共に、主記憶データ制御信号中のアドレスを書き込みアドレスとしてセクタ704を通じてキャッシュ71~74へ送出する。

【0062】選択部7022は、入力レジスタ701から入力されているフラグF1~F4のうち、論理“1”となっているフラグが1以上存在すれば、論理“1”であるフラグの1つを選択し、その選択したフラグに対応するキャッシュに対して更新要求11~14を出力する。例えば、フラグF1のみ論理“1”の場合、更新要求11のみを出力する。また、フラグF1、F2の2つが論理“1”の場合、その何れか一方に対応する更新要求11のみ、または更新要求12のみを出力する。全てのフラグF1~F4が論理“0”の場合、何れの更新要求11~14も出力しない。

【0063】図7に選択部7022の構成例を示す。4ビットの循環型のシフトレジスタ70221には、1ビットのみ論理“1”、残り3ビットは全て論理“0”の4ビットのビットパターン（例えば0001）を保持し、オア回路70222の出力が論理“1”になる毎に1ビットだけビットパターンをシフトする。例えば0001を1ビットだけシフトすると、0010となる。1000の次は0001に戻る。シフトレジスタ70221の第1段目のビットの値は更新要求生成部70223

へ、同様に第2段目、第3段目、第4段目のビットの値は更新要求生成部70224、70225、70226へ、それぞれ出力されている。

【0064】更新要求生成部70223~70226には、入力レジスタからのフラグF1~F4が入力されており、制御部7021によって起動されると、この4つの更新要求生成部70223~70226のうち、シフトレジスタ70221からのビット出力が論理"1"になっている唯一の更新要求生成部のみが動作する。各更新要求生成部70223~70226は、フラグF1~F4の値に応じて、図8に示すような更新要求1~4の組み合わせパターンを発生する。

【0065】図8に示されるように、更新要求生成部70223は、フラグF1を最高の優先度、次にフラグF2、次にフラグF3、最低の優先度をフラグF4とし、より高い優先度のフラグが論理"1"のとき、そのフラグに対応する更新要求1~4のみ論理"1"とする組み合わせパターンを発生する。また、更新要求生成部70224は、フラグF2を最高の優先度、次にフラグF3、次にフラグF4、最低の優先度をフラグF1とし、更新要求生成部70225は、フラグF3を最高の優先度、次にフラグF4、次にフラグF1、最低の優先度をフラグF2とし、更新要求生成部70226は、フラグF4を最高の優先度、次にフラグF3、次にフラグF2、最低の優先度をフラグF1とし、それぞれ更新要求生成部70223と同様に、より高い優先度のフラグが論理"1"のとき、そのフラグに対応する更新要求1~4のみ論理"1"とする組み合わせパターンを発生する。なお、フラグF1~F4の全てが論理"0"の場合、更新要求1~4は全て論理"0"となる。

【0066】更新要求生成部70223~70226から出力される更新要求1~4の組み合わせパターンはオア回路70227~70300を通じて各セットのキャッシュ71~74に送出される。また、更新要求1~4の何れか1つが論理"1"のとき、オア回路70222の出力が論理"1"となり、シフトレジスタ70221が1ビットだけシフトされる。

【0067】図7に示した選択部7022によれば、キャッシュ更新が1回行われる毎に、次のキャッシュ更新時に起動する唯一の更新要求生成部70223~70226が順番に切り替わるため、キャッシュ更新対象となるキャッシュ71~74を順に切り替えていくことができる。例えば、キャッシュ71とキャッシュ72を使うプロセスでは、フラグF1、F2のみが論理"1"となるが、そのプロセスの実行中に連続してキャッシュミスが発生すると、キャッシュ71とキャッシュ72とが交互に更新される。

【0068】

【発明の他の実施の形態】図9は本発明の別の実施形態のブロック図である。図1の実施形態と相違するところ

は、命令キャッシュ6とデータキャッシュ7とでキャッシュのセット数が相違する点である。つまり、本実施形態の場合、命令キャッシュ6はキャッシュ61、62の2セットのキャッシュを持ち、データキャッシュ7はキャッシュ71~74の4セットのキャッシュを持っている。この場合、キャッシュ61とキャッシュ71とがセット1のキャッシュ、キャッシュ62とキャッシュ72とがセット2のキャッシュ、キャッシュ73がセット3のキャッシュ、キャッシュ74がセット4のキャッシュを構成する。

【0069】本実施形態の場合、主記憶アクセス命令に付加するキャッシュ選択フラグFは図3と同様に4つのフラグF1~F4で構成することができる。この場合、命令キャッシュ6に関しては、フラグF1~F4の内、フラグF1とフラグF2の2つのフラグだけが使用される。このように一般的には、命令キャッシュ6を構成するキャッシュのセット数をM、データキャッシュ7を構成するキャッシュのセット数をN、MとNの内の大きな方の値をMNとすると、主記憶アクセス命令にはMN個のキャッシュ選択フラグがあれば十分である。そして、命令キャッシュ6をアクセスする場合には、MN個のキャッシュ選択フラグの内の先頭からN個のキャッシュ選択フラグを使用し、データキャッシュ7をアクセスする場合には、MN個のキャッシュ選択フラグの内の先頭からM個のキャッシュ選択フラグを使用すれば良い。

【0070】キャッシュ選択フラグFの他の構成例として、命令キャッシュ用とデータキャッシュ用に別々のフラグを設けることも可能である。一般的に言えば、命令キャッシュ6を構成するキャッシュのセット数をM、データキャッシュ7を構成するキャッシュのセット数をNとすると、主記憶アクセス命令にM個の命令用のキャッシュ選択フラグとN個のデータ用のキャッシュ選択フラグとを設ける。そして、命令キャッシュ6をアクセスする場合は命令用のキャッシュ選択フラグを使用し、データキャッシュ7をアクセスする場合はデータ用のキャッシュ選択フラグを使用する。例えば図9の実施形態では、図10に示すように、命令キャッシュ6用として2つのフラグF21、F22を設け、データキャッシュ7用として4つのフラグF11~F14を設ける。

【0071】以上の実施形態では、命令キャッシュ6及びデータキャッシュ7の双方を複数のセットのキャッシュで構成したが、何れか一方のみを複数のセットのキャッシュで構成し、他方は1セットのキャッシュで構成したキャッシュメモリに対しても本発明は適用可能である。命令キャッシュ6を1セットのキャッシュで構成した実施形態のブロック図を図11に示す。この場合、1セット構成の命令キャッシュまたはデータキャッシュの当該1つのキャッシュを主記憶アクセス命令中の1つのキャッシュ選択フラグによって使用/不使用を制御しても良いし、キャッシュ選択フラグによる制御は行わず、

常に使用するように制御しても良い。つまり、主記憶アクセス命令によって選択するキャッシュを複数セット構成の命令キャッシュまたはデータキャッシュに限定しても良い。

【0072】更に、以上の実施例では、命令用とデータ用に別々のキャッシュを持つキャッシュメモリに本発明を適用したが、命令用とデータ用と同じ複数のキャッシュを使うキャッシュメモリに対しても本発明は適用可能である。図12に命令用とデータ用とで兼用する4つのセットのキャッシュ61'～64'とそれを制御するキャッシュコントローラ65'を含む実施の形態のブロック図を示す。

【0073】なお、複数のプロセスが異なるセットのキャッシュを使う場合、キャッシュコヒーレントの問題が生じる。例えば図13に示すように、プロセスP1がキャッシュAを使用し、プロセスP2がキャッシュBを使用しており、プロセスP1、P2は主記憶のデータDを共用しているとする。この場合、例えばプロセスP1がキャッシュA上でデータDをD'に更新すると、主記憶上のデータDはD'に更新されるが、キャッシュB上のデータDは更新されない。このため、プロセスP2がキャッシュB上でデータDを参照すると、古いデータを参照することになる。このようなキャッシュコヒーレントの問題を解消する手法には、ソフトウェア的に解消する方法とハードウェア的に解消する方法とがあり、本発明はその何れを使用しても良い。ソフトウェア的な解消方法としては、各プロセスが異なるキャッシュセットを使う他プロセスと共用するデータを参照するときに、自プロセスで使うキャッシュの内容をクリアしてから参照する方法などがある。また、ハードウェア的な解消方法としては、各セットのキャッシュに、キャッシュ更新時に更新アドレスを他の全てのキャッシュに報告する手段と、他キャッシュから前記報告を受けたときに若し自キャッシュにその更新アドレスがキャッシュされていれば、それを無効化する手段とを設ける方法などがある。

【0074】

【発明の効果】以上説明したように本発明によれば、複数のキャッシュを備えたキャッシュメモリにおいて、複数のキャッシュの中から実際に使用する任意個数のキャッシュを、主記憶アクセス命令によって直接選択することができる。その理由は、キャッシュのセットに1対1に対応するキャッシュ選択フラグの状態に応じてキャッシュのセットを選択するため、複数のキャッシュのセットの中から1つのセットを選択する以外に、2つ以上のセットを選択したり、全てのセットを選択することもでき、またセットを1つも選択しないこともできるからである。

【0075】また本発明によれば、任意数のキャッシュセットの選択が主記憶アクセス命令毎に行えるため、プ

ロセス毎に、そのプロセスで使用する主記憶アクセス命令中のキャッシュ選択フラグを調整することで、プロセス単位でのキャッシュ選択が可能になり、更に実行中のプロセスに通じた容量のキャッシュを用いるといったキャッシュ構成をオーバーヘッド無しに実施することが可能となる。

【0076】以上のような効果を実現するため、複数のプロセスを並行して実行する計算機システムに本発明を適用することにより、計算機システムの性能向上が可能となる。特に並列プロセス数が比較的少なく高速な処理が要求されるスーパーコンピュータ等の大規模科学技術計算を行う計算機に適用すれば、より効果的である。

【図面の簡単な説明】

【図1】本発明を適用した計算機システムの一例を示すブロック図である。

【図2】本発明で使用する主記憶アクセス命令のうち、一般的なロード命令とストア命令についてのフォーマット例を示す図である。

【図3】キャッシュ選択フラグの内容例を示す図である。

【図4】本発明の一実施形態における一実施例のブロック図である。

【図5】各セットのキャッシュの構成例を示すブロック図である。

【図6】キャッシュ内の更新制御部の構成例を示すブロック図である。

【図7】更新制御部内の選択部の構成例を示すブロック図である。

【図8】選択部内の更新要求生成部の入力と出力との関係を示す図である。

【図9】本発明の別の実施形態のブロック図である。

【図10】キャッシュ選択フラグの別の内容例を示す図である。

【図11】本発明の更に別の実施形態のブロック図である。

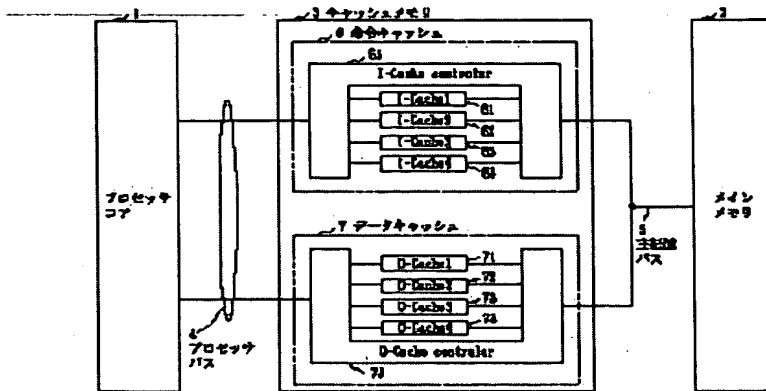
【図12】本発明の他の実施形態のブロック図である。

【図13】キャッシュコヒーレント問題の説明図である。

【符号の説明】

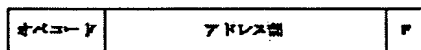
- 1…プロセッサコア
- 2…主記憶（メインメモリ）
- 3…キャッシュメモリ
- 4…プロセッサバス
- 5…主記憶バス
- 6…命令キャッシュ
- 7…データキャッシュ
- 61～64、71～74…キャッシュ
- 65、75…キャッシュコントローラ

【図1】

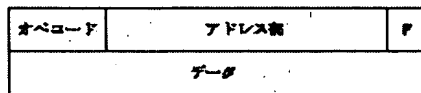


【図2】

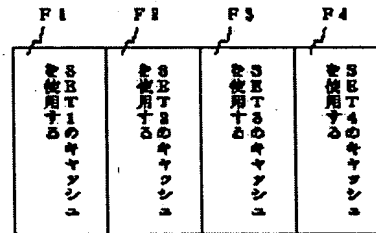
Load命令



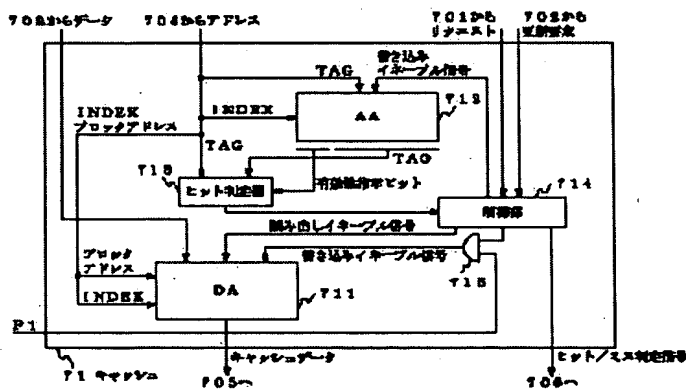
Store命令



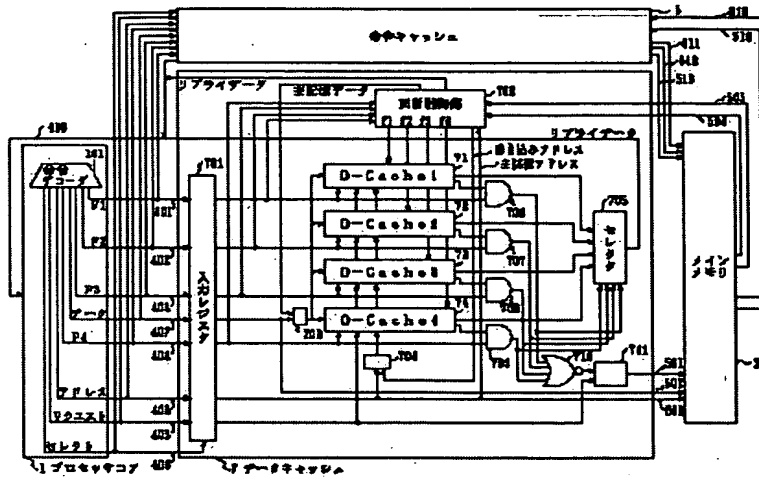
【図3】



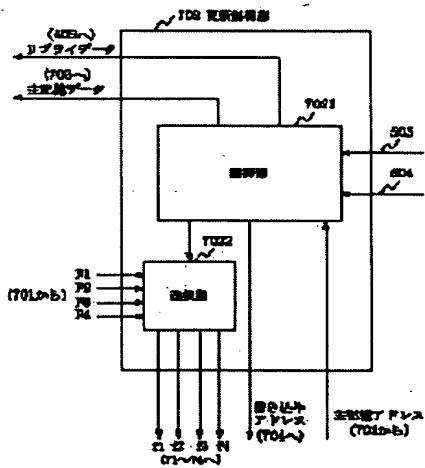
【図5】



【図4】



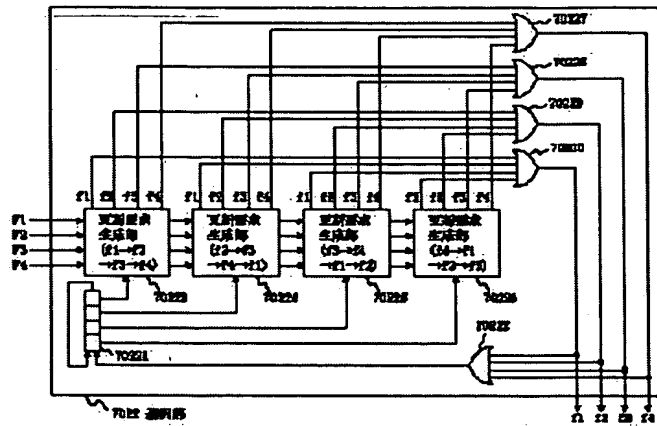
【図6】



【図10】

P11	P12	P13	P14	P31	P32
ディタマシンを 使用する	ディタマシンを 使用する	ディタマシンを 使用する	ディタマシンを 使用する	ディタマシンを 使用する	ディタマシンを 使用する

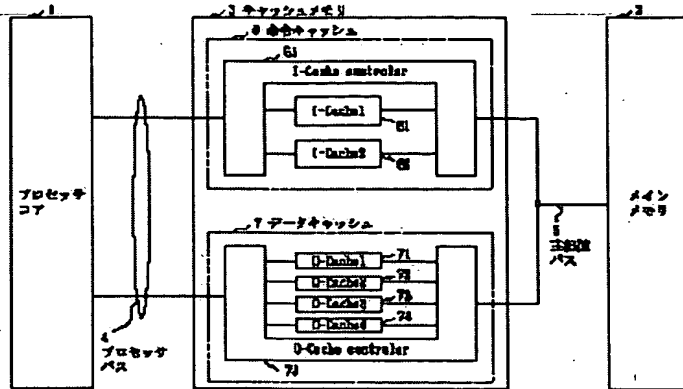
【图7】



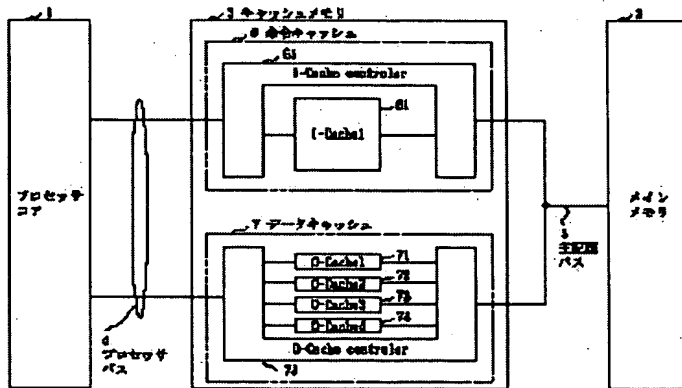
【图8】

F1 F2 F3 F4	74221 4位并行加法器	74222 4位并行加法器	74223 4位并行加法器	74224 4位并行加法器
0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0
0 0 0 1	0 0 0 1	0 0 0 1	0 0 0 1	0 0 0 1
0 0 1 0	0 0 1 0	0 0 1 0	0 0 1 0	0 0 1 0
0 0 1 1	0 0 1 1	0 0 1 1	0 0 1 1	0 0 1 1
0 1 0 0	0 1 0 0	0 1 0 0	0 1 0 0	0 1 0 0
0 1 0 1	0 1 0 1	0 1 0 1	0 1 0 1	0 1 0 1
0 1 1 0	0 1 1 0	0 1 1 0	0 1 1 0	0 1 1 0
0 1 1 1	0 1 1 1	0 1 1 1	0 1 1 1	0 1 1 1
1 0 0 0	1 0 0 0	1 0 0 0	1 0 0 0	1 0 0 0
1 0 0 1	1 0 0 1	1 0 0 1	1 0 0 1	1 0 0 1
1 0 1 0	1 0 1 0	1 0 1 0	1 0 1 0	1 0 1 0
1 0 1 1	1 0 1 1	1 0 1 1	1 0 1 1	1 0 1 1
1 1 0 0	1 1 0 0	1 1 0 0	1 1 0 0	1 1 0 0
1 1 0 1	1 1 0 1	1 1 0 1	1 1 0 1	1 1 0 1
1 1 1 0	1 1 1 0	1 1 1 0	1 1 1 0	1 1 1 0
1 1 1 1	1 1 1 1	1 1 1 1	1 1 1 1	1 1 1 1

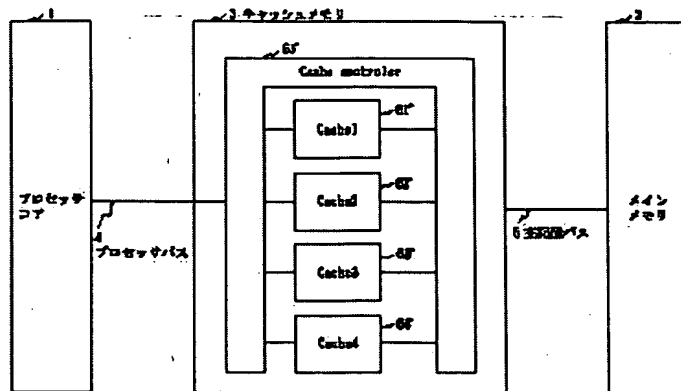
【図9】



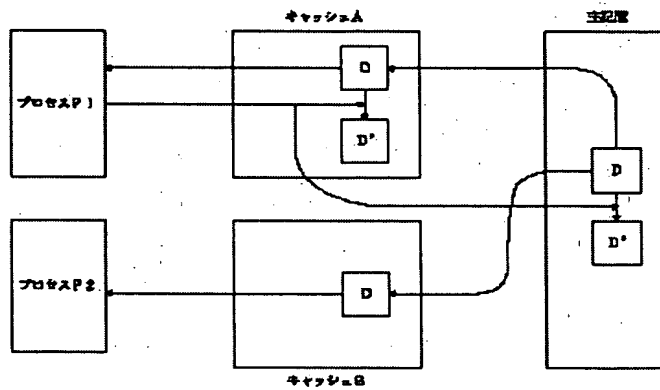
【図11】



【図12】



【図13】



フロントページの続き

(51) Int. Cl. 7	識別記号	FI	テーマコード* (参考)
G06F 9/34	350	G06F 9/34	350B
9/46	340	9/46	340F